

Contents

1	Introduction.	1
	References	8
2	A Primer in Genetics	9
2.1	Basic Biology	9
2.1.1	Phenotypes and Genotypes	9
2.1.2	Meiosis and Crossover	12
2.1.3	Genetic Distance	12
2.1.4	The Haldane Mapping Function.	13
2.1.5	Interference and Other Mapping Functions	14
2.1.6	Markers and Genetic Maps	16
2.2	Types of Study	17
2.2.1	Crossing Experiments.	17
2.2.2	The Basics of QTL Mapping.	21
2.2.3	Association Studies	23
2.2.4	Other Types of Study.	27
	References	29
3	Statistical Methods in High Dimensions.	31
3.1	Overview	31
3.2	Multiple Testing	32
3.2.1	Classical Procedures Controlling FWER.	33
3.2.2	Permutation Tests and Resampling Procedures	37
3.2.3	Controlling the False Discovery Rate	41
3.2.4	Multiple Testing Under Sparsity. Minimizing the Bayesian Risk in Multiple Testing Procedures.	42
3.3	Model Selection.	51
3.3.1	The Likelihood Function	52
3.3.2	Information Theoretical Approach	55
3.3.3	Bayesian Model Selection and the Bayesian Information Criterion	57

3.3.4	Modifications of BIC for High-Dimensional Data Under Sparsity	59
3.3.5	Further Approaches to Model Selection	61
	References	69
4	Statistical Methods of QTL Mapping for Experimental Populations	73
4.1	Classical Approaches	73
4.1.1	Single Marker Tests	73
4.1.2	Power of a Test Based on a Single Marker as a Function of the Distance Between the Marker and a QTL	74
4.1.3	Genome Wide Search with Tests Based on Single Markers	76
4.2	Interval Mapping	79
4.2.1	Interval Mapping Based on the Mixture Model	79
4.2.2	Regression Interval Mapping	81
4.2.3	Nonparametric Version of Interval Mapping	82
4.2.4	Specific Models	83
4.2.5	Overestimation of Genetic Effects	83
4.3	Model Selection	84
4.3.1	QTL mapping with mBIC	84
4.3.2	Robust Version of mBIC	87
4.3.3	Version of mBIC Based on Rank Regression	89
4.3.4	Extensions to Generalized Linear Models	90
4.3.5	mBIC for Dense Markers and Interval Mapping	92
4.4	Logic Regression	97
4.5	Applying mBIC in a Bayesian Approach	101
4.6	Closing Remarks	101
	References	102
5	Statistical Analysis of GWAS	105
5.1	Overview	105
5.2	Inferring Genotypes	107
5.2.1	Genotype Calling	107
5.2.2	Imputation	110
5.3	Single Marker Tests	111
5.3.1	Case-Control Studies	111
5.3.2	Quantitative Traits	115
5.3.3	Covariates and Population Stratification	118
5.3.4	Multiple Testing Correction	121
5.3.5	Rare SNPs	122
5.4	Model Selection	124
5.4.1	Motivation	124
5.4.2	HYPERLASSO	129

5.4.3	GWASelect.	132
5.4.4	MOSGWA	133
5.4.5	Comparison of Methods	135
5.4.6	Mixed Models.	138
5.5	Admixture Mapping	140
5.6	Gene–Gene Interaction	144
5.6.1	Analyzing Gene–Gene Interaction via ANOVA	144
5.6.2	Multifactor Dimensionality Reduction	147
5.6.3	Logic Regression in GWAS	149
5.7	Other Recent Advances and the Outlook for GWAS.	151
	References	156
6	Appendix A: Basic Statistical Distributions	163
6.1	Normal Distribution	163
6.2	Important Distributions of Sample Statistics	165
6.2.1	Chi-Square Distribution	165
6.2.2	Student’s t-Distribution.	166
6.2.3	F-distribution	167
6.3	Gamma and Beta Distributions	168
6.3.1	Exponential Distribution	168
6.3.2	Inverse Gamma Distribution	168
6.3.3	Beta Distribution	169
6.4	Double Exponential Distribution and Extensions	169
6.4.1	Asymmetric Double Exponential (ADE) Distribution.	169
6.5	Discrete Distributions	170
6.5.1	Binomial Distribution.	170
6.5.2	Poisson Distribution.	170
6.5.3	Negative Binomial Distribution	171
6.5.4	Generalized Poisson Distribution	171
6.5.5	Zero-Inflated Generalized Poisson Distribution	172
	Reference	172
7	Appendix B: Basic Methods of Estimation.	173
7.1	Basic Properties of Statistical Estimators.	173
7.1.1	Statistical Bias.	173
7.1.2	Mean Square Error	174
7.1.3	Efficiency of Estimators	174
7.1.4	Method of Moments	175
7.1.5	Maximum Likelihood Estimation.	176
7.2	Estimates of Basic Statistical Parameters.	177
7.2.1	Mean and Variance	177
7.2.2	Pearson Correlation Coefficient	177
	Reference	178

8	Appendix C: Principles of Statistical Testing	179
8.1	Basic Ideas of Statistical Testing: The Z-test	179
8.2	The Family of t-tests	182
8.2.1	One Sample t-Test	182
8.2.2	Two Sample t-Test	183
8.2.3	Paired t-Test	184
8.2.4	Robustness of t-Tests	184
8.3	Classical Approach to ANOVA and Regression	185
8.3.1	One-Way Analysis of Variance	185
8.3.2	Two-Way ANOVA. Interactions	186
8.3.3	Two-Way ANOVA with No Interactions	189
8.3.4	Extensions to a Larger Number of Factors	190
8.3.5	Multiple Regression	190
8.3.6	Weighted Least Squares	192
8.4	General Linear Models	192
8.4.1	Cockerham's Model	193
8.4.2	Robustness of General Linear Models	194
8.5	Generalized Linear Models	194
8.5.1	Extensions of Poisson Regression	197
8.6	Linear Mixed Models	197
8.7	Nonparametric Tests	200
8.7.1	Wilcoxon Signed-Rank Test	202
8.7.2	Rank Regression	202
8.8	Tests for Categorical Variables	203
8.8.1	Chi-Square Goodness-of-Fit Test	203
8.8.2	Chi-Square Test of Independence	204
8.8.3	Fisher's Exact Test	205
	References	206
9	Appendix D: Elements of Bayesian Statistics	207
9.1	Bayes Rule	207
9.2	Conjugate Priors	208
9.3	Markov Chain Monte Carlo	208
9.3.1	Gibbs Sampler	209
9.3.2	Metropolis–Hastings Algorithm	210
9.3.3	Hierarchical Models	210
9.3.4	Parametric Empirical Bayes	211
9.4	Bayes Classifier	212
	References	212
10	Appendix E: Other Statistical Methods	215
10.1	Principal Component Analysis	215
10.2	The EM Algorithm	216
	References	217
	Index	219

Phenotypes and Genotypes

The Search for Influential Genes

Frommlet, F.; Bogdan, M.; Ramsey, D.

2016, XII, 222 p. 33 illus., 20 illus. in color., Hardcover

ISBN: 978-1-4471-5309-2